

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 13:24:37 ; Search time 3099 Seconds
(without alignments)
6055.400 Million cell updates/sec

Title: US-10-071-510a-16
Perfect score: 493
Sequence: 1 cggcggccgggggagtcg.....ccctgtctcatttgagcctgc 493

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_hcc.*
- 4: gb_est3.*
- 5: gb_est4.*
- 6: gb_est5.*
- 7: gb_est6.*
- 8: gb_g881.*
- 9: gb_g882.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	212.2	43.0	5069	AK029216	Mus musculus
2	180.6	38.3	3080	AK030542	Mus musculus
3	177.6	36.0	384	AQ033373	HS-2227 A
4	163.2	33.1	248	AW357996	41821 MAR
5	121.4	24.6	438	AZ971121	2M0244E01
6	117.6	23.9	596	CN718887	E0774B03-
7	109.2	22.2	368	BY014727	BY014727
8	85.6	17.4	822	CN227139	RJB088A07
9	55.2	11.2	1688	CR709987	Tetraodon
10	53.4	10.8	695	CD574999	UCRPT01 0
11	52.4	10.6	635	RJ312765	BJ312765
12	52.4	10.6	692	CD878716	CD878716
13	51.8	10.5	1176	CL974133	CL974133
14	51.2	10.4	413	BQ980809	BQ980809
15	51	10.3	709	AG351718	AG351718
16	51	10.3	776	AG561854	AG561854
17	50.6	10.3	632	BM324787	BM324787
18	49.8	10.1	1356	CL945698	CL945698
19	49.8	10.1	1411	CL641584	CL641584
20	49.4	10.0	426	CK273503	CK273503
21	49.2	10.0	557	BU010846	BU010846
22	49.2	10.0	565	BQ983268	BQ983268
23	49	9.9	242	BQ990877	BQ990877
24	48.6	9.9	591	CV429907	CV429907

C	25	48.6	9.9	691	7	CN604864
C	26	48.6	9.9	810	7	CK204412
C	27	48.4	9.8	400	6	CD575000
C	28	48.2	9.8	428	1	AU196028
C	29	48.2	9.8	502	1	AU195622
C	30	48.2	9.8	513	1	AU190638
C	31	48.2	9.8	525	1	AV429680
C	32	48.2	9.8	534	1	AU191794
C	33	48.2	9.8	538	5	BQ982020
C	34	48.2	9.8	542	1	AU195424
C	35	48	9.7	347	7	CN604887
C	36	48	9.7	501	7	CV094178
C	37	48	9.7	520	7	CV096855
C	38	48	9.7	552	7	CN605345
C	39	48	9.7	592	7	CV092833
C	40	48	9.7	609	7	CV099895
C	41	48	9.7	624	7	CV092434
C	42	48	9.7	647	7	CV096641
C	43	48	9.7	651	7	CV093263
C	44	48	9.7	651	7	CV097413
C	45	48	9.7	752	6	CB649785

ALIGNMENTS

RESULT 1	AK029216	5069 bp	mrna	linear	HTC 03-APR-2004
LOCUS	Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched library, clone:4831426119 product:hypothetical Spectrin repeat containing protein, full insert sequence.				
DEFINITION	AK029216	1	GI:26325193		
ACCESSION	AK029216				
VERSION	HTC; CAP trapper.				
KEYWORDS	Mus musculus (house mouse)				
SOURCE	Mus musculus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1	Carninci, P. and Hayashizaki, Y.			
AUTHORS	High-efficiency full-length cDNA cloning				
TITLE	Meth. Enzymol. 303, 19-44 (1999)				
JOURNAL	99279253				
MEDLINE	10349636				
PUBMED					
REFERENCE	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
AUTHORS	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
TITLE	Genome Res. 10 (10), 1617-1630 (2000)				
JOURNAL	20499374				
MEDLINE	11042159				
PUBMED					
REFERENCE	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, K., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.			
AUTHORS	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer				
TITLE	Genome Res. 10 (11), 1757-1771 (2000)				
JOURNAL	20530913				
MEDLINE	11076861				
PUBMED					
REFERENCE	4	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.			
AUTHORS	Functional annotation of a full-length mouse cDNA collection				
TITLE	Nature 409, 685-690 (2001)				
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE	5	The FANTOM Consortium and the RIKEN Genome Exploration Research			
AUTHORS					

Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
Of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 5069)
ADACHI, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Nomazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saito, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.
URL: http://fantom.gsc.riken.jp/.
FEATURES
source
1. 5069
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:4831426119"
/db_xref="taxon:10090"
/clone="4831426119"
/tissue types="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
201..2888
/notes="unnamed protein product; hypothetical Spectrin
repeat containing protein (InterPro|IPR002017, evidence:
InterPro)
putative"
/codon_start=1
/protein_id="BAC26351.1"
/db_xref="GI:26325194"
/translation="MTTRGPERPWRQGFERSQRIEWMVLMHSEYLLAQDFYRWFQK
MVVALPPVELQLGKEKQWLSHAQVLLHNVNDQAVLLDRLLLEAGSLFSRIGDPDV
DEDAQKRMKAEYDAVKARQRRVDLLAQVAQHEOYREDVNEFQMLKAVKVSCL
GRNCKLATELRTLIQDIADKPPRGESLKLAEQAVGVIQNTSPLEGKISGELSEEM
GVLEKRLVWKEEERKGLQSDCEQIQQLAEGLGDFPKISLQRLAQEGLEETV
KTATEDLVNQWLFSGTAALESFPRVDRLQTLKLVTFPDLQSLDSVVVATIEP
YQSMGKNTLRLNATRAELWQRFPLNDLQMLKALRLDITASLPDLASTHTFLP
QIEAALTESRLKEQLMLQKLDLGSIFGQRAATLLLEQVTSVDRDLRLHNSLQ
RKSLOSILVQHKDFGVAFDPLNRKLLDQARIQAEKGLPRDLPGKQVOLLRLQGLQE
EGLDLQTEAVRPLAHGNSKQKQVDQISCDQQAQLRSLDLVRCQNVREHCTFS
HRUSELQIMTATQLESQHVDRWDAESQAGLETLLSEIPEKVEQVSLQALQGQ
LVNKKSPGATVQELKIMESQWALRLLEENLSLMRNQQLQRTVEDTGGKQVFT
NNIPKAGFIINPDQPIPRQGANPLEGHDLPEDHQLLRQAGPLQAGRSALPDVAVQV
MRVATKDLTRVQLQLEARIPEGCHLPENLRQAGKPLQAGRSALPDVAVQV
QAQGLPPADRESFGADCTPESAEAVESLSPTESMPCGTAAIAVAPALSAAVAP
AAGRRGAQLRPQQRLPRLCAHASVQWPAHLAHWAHR"

7 GCCCGGGGATCCCGAGTCCCAAGAGCGCGAGTTTGTAGAGGCTGGTGGCAGAAATCCCCG 66
1972 GTCTGTGGGATGCTAGTCTCCCAAGAGCGTGGAGCTCGAGAGCTGTCTGTCTGAAATCCGAG 2031
67 AGAAGGAGGCCAGCTGTCTCCCTGGTGAAGCAGCAGGCTGGTGGTGTGATGGAGAACTCTT 126
2032 AGAAAGAGGTCTCAGGTGTCTCCCTGCTCCAAGCACTGGGCCAGCTTGTGTATGAAGAAGCTCT 2091
127 CTCGGAGAGGTCTGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 186
2092 CCCCAAGAGGGCAACCATCGTGGTCCAGGAGGAGCTGAGGAAGCTGTGATGGAGTCTTTGGCAGG 2151
187 CTTTGAGGCTGTGGAAGAAGTCTGCTCAGCGCTCATCAGAAACTGGCATCTGCAGAGGA 246
2152 CCTCGGGCTGTAGAGGAACATGCTGAGTCTCATGAGAAACCAACAGCAGCTGCAGAGGA 2211
247 TGGAAAGTGATTCGGGGAGAAATAGTTTTCACCAACCAATCCCAAGTCAGGATTTTC 306
2212 CAGAGTGGACACGGGGAAGAGCAGGTGTTCACCAACAACATCCCAAGGCGCGCTTTC 2271
307 TCATCAATCCATGATCTCTATTCCTCCAGG 335
2272 TCATCAACCTCAGGACCCCATTCCTCCAGG 2300

RESULT 2
AK030542
LOCUS
DEFINITION
AK030542 3080 bp mRNA linear HTC 03-APR-2004
Mus musculus adult male pituitary gland cDNA, RIKEN full-length
enriched library, clone:5330429E21 product:hypothetical Spectrin
repeat containing protein, full insert sequence.
ACCESSION
VERSION AK030542.1 GI:26326536
KEYWORDS
SOURCE HTC; CAP trapper.
ORGANISM Mus musculus (house mouse)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED 10349636
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED 11042159
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED 11076861
4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
REFERENCE
5 The FANTOM Consortium and the RIKEN Genome Exploration Research

Query Match 43.0%; Score 212.2; DB 3; Length 5069;
Best Local Similarity 77.8%; Pred. No. 2.8e-42;
Matches 256; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

ORIGIN

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 19:37:01 ; Search time 420 Seconds
(without alignments)

6756.145 Million cell updates/sec

Title: US-10-071-510A-16

Perfect score: 493

Sequence: 1 cggccggccggggatgcc.....cctgtctcattgagcctgc 493

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4313806 seqs, 2877871033 residues

Total number of hits satisfying chosen parameters: 8627612

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US10F_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	493	100.0	493	15	US-10-071-510-16
C 2	176.4	35.8	532	16	US-10-029-386-10669
C 3	174	35.3	174	16	US-10-029-386-24369
C 4	57	11.6	6302	18	US-10-437-963-93121
C 5	56.8	11.5	2282	18	US-10-479-638-7
C 6	56	11.4	6402	18	US-10-437-963-16557
C 7	53.6	10.9	6296	18	US-10-437-963-16601
C 8	53	10.8	458	18	US-10-437-963-3998
C 9	53	10.8	1023	18	US-10-437-963-56215
C 10	53	10.8	6112	18	US-10-437-963-93061
C 11	53	10.8	6417	18	US-10-437-963-16524

C 12	52.4	10.6	6337	18	US-10-437-963-16612	Sequence 16612, A
C 13	50.6	10.3	2139	18	US-10-437-963-69317	Sequence 69317, A
C 14	50.6	10.3	4701	18	US-10-437-963-69371	Sequence 69371, A
C 15	50.6	10.3	6972	18	US-10-437-963-16522	Sequence 16522, A
C 16	50.2	10.2	1612	18	US-10-437-963-102480	Sequence 102480, A
C 17	50.2	10.2	15231	9	US-09-917-800A-1505	Sequence 1505, Ap
C 18	50.2	10.2	15231	17	US-10-152-319A-1780	Sequence 1780, Ap
C 19	49.8	10.1	768	9	US-09-938-842A-812	Sequence 812, App
C 20	49.8	10.1	768	11	US-09-938-842A-812	Sequence 812, App
C 21	49.8	10.1	2607	18	US-10-437-963-69298	Sequence 69298, A
C 22	49.8	10.1	2689	18	US-10-437-963-69366	Sequence 69366, A
C 23	49.8	10.1	4524	18	US-10-437-963-41480	Sequence 41480, A
C 24	49.8	10.1	16442	17	US-10-374-077-288	Sequence 208, App
C 25	49	9.9	860	18	US-10-767-701-3533	Sequence 3533, Ap
C 26	48.6	9.9	2853	18	US-10-437-963-69296	Sequence 69296, A
C 27	48.6	9.9	5823	18	US-10-437-963-69258	Sequence 69258, A
C 28	48.6	9.9	114793	15	US-10-148-806-3	Sequence 3, Appl1
C 29	48.2	9.8	628	16	US-10-029-386-22859	Sequence 22859, A
C 30	48.2	9.8	5738	18	US-10-437-963-93231	Sequence 93231, A
C 31	48.2	9.8	6332	18	US-10-437-963-16484	Sequence 16484, A
C 32	47.8	9.7	1926	15	US-10-294-804-3	Sequence 3, Appl1
C 33	47.8	9.7	97	18	US-10-194-046-3	Sequence 76755, A
C 34	47.8	9.7	2820	18	US-10-437-963-76755	Sequence 93087, A
C 35	47.8	9.7	6400	18	US-10-437-963-93087	Sequence 14, Appl
C 36	47.8	9.7	8705	15	US-10-291-230-14	Sequence 16, Appl
C 37	47.8	9.7	8705	17	US-10-273-678-16	Sequence 1, Appl1
C 38	47.8	9.7	97	18	US-10-273-751-1	Sequence 283, App
C 39	47.8	9.7	9600	16	US-10-050-898-283	Sequence 283, App
C 40	47.8	9.7	10233	16	US-10-050-902-283	Sequence 24, Appl
C 41	47.8	9.7	10285	16	US-10-656-269-24	Sequence 22, Appl
C 42	47.8	9.7	10330	18	US-10-656-269-22	Sequence 20, Appl
C 43	47.8	9.7	10477	18	US-10-656-269-20	Sequence 18, Appl
C 44	47.8	9.7	10516	18	US-10-656-269-20	
C 45	47.8	9.7	10561	18	US-10-656-269-18	

ALIGNMENTS

RESULT 1

US-10-071-510-16

; Sequence 16, Application US/10071510

; Publication No. US20030143552A1

; GENERAL INFORMATION:

; APPLICANT: Clark, Edwin

; APPLICANT: Grenfell, Tallesyn

; APPLICANT: Lu, Karen

; APPLICANT: Hartmann, Lynn

; APPLICANT: Brown, Jeffrey L.

; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR THE

; IDENTIFICATION, ASSESSMENT, PREVENTION AND THERAPY

; TITLE OF INVENTION: OF HUMAN CANCERS

; FILE REFERENCE: MRI-027

; CURRENT APPLICATION NUMBER: US/10/071,510

; CURRENT FILING DATE: 2002-02-08

; PRIOR APPLICATION NUMBER: US 60/267,276

; PRIOR FILING DATE: 2001-02-08

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 16

; LENGTH: 493

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-071-510-16

Query Match 100.0%; Score 493; DB 15; Length 493;

Best Local Similarity 100.0%; Pred. No. 4.2e-137;

Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

1 CGGCGCGCGCGGATGCCAGTCCCAAGAGGCCGAGTTTGAGAGGCTGTGGCAGAAAT 60

Db

1 CGGCGCGCGCGGATGCCAGTCCCAAGAGGCCGAGTTTGAGAGGCTGTGGCAGAAAT 60

QY 61 TCCCGAAGAGGAGCCAGCTGTCCCTGGTGAAGCGCAGGCGCTGGCTGGTGAAGAG 120
Db 61 TCCCGAAGAGGAGCCAGCTGTCCCTGGTGAAGCGCAGGCGCTGGCTGGTGAAGAG 120
QY 121 AGTCTTCTCCGAGGAGTGTCCGCTGGTGCAGAGAGCTCAGGAGCTGGCAGAGTCGT 180
Db 121 AGTCTTCTCCGAGGAGTGTCCGCTGGTGCAGAGAGCTCAGGAGCTGGCAGAGTCGT 180
QY 181 GCGGCGCTTGGCTGTGGAGAAGTCTGTGAGCGCTCATCAAACTGGCATCTGC 240
Db 181 GCGGCGCTTGGCTGTGGAGAAGTCTGTGAGCGCTCATCAAACTGGCATCTGC 240
QY 241 AGAGGATGGAAGTGGATTCGGGGAAGAAATGGTTTTTCCAAACAATCCCAAAGTCAG 300
Db 241 AGAGGATGGAAGTGGATTCGGGGAAGAAATGGTTTTTCCAAACAATCCCAAAGTCAG 300
QY 301 GATTTCTCATCAATCCCATGATTCCTATTTCCAGGCAATCGTCAGCGTGAGTCTCTTA 360
Db 301 GATTTCTCATCAATCCCATGATTCCTATTTCCAGGCAATCGTCAGCGTGAGTCTCTTA 360
QY 361 GCAGGCGCTGGGAGAAGGGCCAGCGCCAGCTCAAGAGTGGGTAGGGGTCTCCAGCA 420
Db 361 GCAGGCGCTGGGAGAAGGGCCAGCGCCAGCTCAAGAGTGGGTAGGGGTCTCCAGCA 420
QY 421 CAGGCGCTCCCTGTCTGGGGCAACATGCTCTGCTGAGGACTTGGCCACGTCCTGTCT 480
Db 421 CAGGCGCTCCCTGTCTGGGGCAACATGCTCTGCTGAGGACTTGGCCACGTCCTGTCT 480
QY 481 CATTTGAGCGCTGC 493
Db 481 CATTTGAGCGCTGC 493

RESULT 2
US-10-029-386-10669/c
; Sequence 10669, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10669
; LENGTH: 532
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR14.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P10476, EVALUE 5.70e-01
; OTHER INFORMATION: NT HIT: AJ277661.1, EVALUE 5.10e-01
; OTHER INFORMATION: EST_HUMAN HIT: AW842384.1, EVALUE 3.00e-03
US-10-029-386-10669

Query Match 35.8%; Score 176.4; DB 16; Length 532;
Best Local Similarity 99.4%; Pred. No. 1.9e-42;
Matches 177; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 40 TTGAGAGGCTGGTGGCAGAAATTCCTGGAGAGGAGGAGCCAGCTGTCCCTGGTGGAGAGCG 99
Db 425 TTGAGAGGCTGGTGGCAGAAATTCCTGGAGAGGAGGAGCCAGCTGTCCCTGGTGGAGAGCG 366
QY 100 AGGCGTGGCTGGTGGAGAAATCTTCTCCGAGGAGTGTCCCGTGGTGCAGGAGGAGC 159
Db 365 AGGCGTGGCTGGTGGAGAAATCTTCTCCGAGGAGTGTCCCGTGGTGCAGGAGGAGC 306

QY 160 TCAGGAGCTGCAGAGTCTGTGCGGGCTTTCAGGCTGTGGAAGAAAGTCTGCTGAG 217
Db 305 TCAGGAGCTGCAGAGTCTGTGCGGGCTTTCAGGCTGTGGAAGAAAGTCTGCTGAG 248
RESULT 3
US-10-029-386-24369/c
; Sequence 24369, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24369
; LENGTH: 174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR14.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: Q99372, EVALUE 6.60e-01
; OTHER INFORMATION: EST_HUMAN HIT: B1255232.1, EVALUE 5.70e-02
US-10-029-386-24369

Query Match 35.3%; Score 174; DB 16; Length 174;
Best Local Similarity 100.0%; Pred. No. 7.4e-42;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 44 GAGGCTGGTGGCAGAAATTCCTGGAGAGGAGCCAGCTGTCCCTGGTGGAGAGCGCAGG 103
Db 174 GAGGCTGGTGGCAGAAATTCCTGGAGAGGAGCCAGCTGTCCCTGGTGGAGAGCGCAGG 115
QY 104 CTGGCTGGTGGTGGAGAAATCTTCTCCGAGGAGTGTCCCGTGGTGCAGGAGAGTCTAG 163
Db 114 CTGGCTGGTGGTGGAGAAATCTTCTCCGAGGAGTGTCCCGTGGTGCAGGAGAGTCTAG 55
QY 164 GCAGCTGGCAGAGTCTGTGCGGGCTTTCAGGCTGTGGAAGAAAGTCTGCTGAG 217
Db 54 GCAGCTGGCAGAGTCTGTGCGGGCTTTCAGGCTGTGGAAGAAAGTCTGCTGAG 1

RESULT 4
US-10-437-963-93121/c
; Sequence 93121, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 93121
; LENGTH: 6302
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 13:24:37 ; Search time 143 Seconds
(without alignments)
5641.150 Million cell updates/sec

Title: US-10-071-510A-16

Perfect score: 493

Sequence: 1 cggcgccggcggggagtcg.....cctgtctcatttgagcctgc 493

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/5A COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51.4	10.4	1505	1	Sequence 1, Appli
2	49.8	10.1	16442	3	Sequence 208, App
3	49.8	10.1	16442	4	Sequence 208, App
4	48.6	9.9	114793	4	Sequence 3, Appli
5	47.8	9.7	913	1	Sequence 3, Appli
6	47.8	9.7	913	1	Sequence 3, Appli
7	47.8	9.7	913	1	Sequence 3, Appli
8	47.8	9.7	913	1	Sequence 3, Appli
9	47.8	9.7	913	1	Sequence 3, Appli
10	47.8	9.7	913	1	Sequence 3, Appli
11	47.8	9.7	1926	3	Sequence 2, Appli
12	47.8	9.7	1926	4	Sequence 25, Appli
13	47.8	9.7	1984	1	Sequence 25, Appli
14	47.8	9.7	1985	1	Sequence 25, Appli
15	47.8	9.7	1985	1	Sequence 25, Appli
16	47.8	9.7	2580	3	Sequence 2, Appli
17	47.8	9.7	2580	3	Sequence 2, Appli
18	47.8	9.7	5452	2	Sequence 1, Appli
19	47.8	9.7	8705	4	Sequence 14, Appli
20	47.8	9.7	9600	3	Sequence 1, Appli
21	47.8	9.7	9600	3	Sequence 1, Appli
22	47.8	9.7	10596	1	Sequence 15, Appli
23	47.8	9.7	10596	1	Sequence 15, Appli
24	47.8	9.7	10596	1	Sequence 15, Appli
25	47.8	9.7	10596	1	Sequence 15, Appli
26	47.8	9.7	10596	2	Sequence 15, Appli
27	47.8	9.7	10596	5	PCT-US93-04648-15

Sequence 48, Appli
Sequence 48, Appli
Sequence 16775, A
Sequence 209, App
Sequence 209, App
Sequence 14, Appli
Sequence 7, Appli
Sequence 80, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 20, Appli
Sequence 20, Appli
Sequence 13845, A

Sequence 48, Appli
Sequence 48, Appli
Sequence 16775, A
Sequence 209, App
Sequence 209, App
Sequence 14, Appli
Sequence 7, Appli
Sequence 80, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 20, Appli
Sequence 20, Appli
Sequence 13845, A

ALIGNMENTS

RESULT 1
US-07-915-246-1
; Sequence 1, Application US/07915246
; Patent No. 5401836
; GENERAL INFORMATION:
; APPLICANT: Baszczynski, Chris L.
; APPLICANT: Fallis, Lynne
; APPLICANT: Bellmare, Guy
; APPLICANT: Boivin, Rodolphe
; TITLE OF INVENTION: A BRASSICA REGULATORY SEQUENCE FOR
; TITLE OF INVENTION: ROOT-SPECIFIC OR ROOT-ABUNDANT GENE EXPRESSION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: McAndrews, Held, and Malloy
; STREET: 500 W. Madison St. 34th Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07915,246
FILING DATE: 19920716
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pochopien, Donald J.
REGISTRATION NUMBER: 32,167
REFERENCE/DOCKET NUMBER: 91 P 1125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312 707-8889
TELEFAX: 312 707-9155

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1505 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Brassica napus
STRAIN: Westar
DEVELOPMENTAL STAGE: Somatic
TISSUE TYPE: Root
US-07-915-246-1

Query Match 10.4%; Score 51.4; DB 1; Length 1505;
 Best Local Similarity 49.8%; Pred. No. 0.00021;
 Matches 130; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 14 GGATCGGAGTCCCAAGAGCCGAGTTTGAGAGGCTGGTGGCAGAAATTCGCCGAGAAGGA 73
 DB 852 GGTATGAGAGTGGAGCCGTGAAGGTGGTGGAGCTGGATACGAGCGCGGAGAGCTGGA 911

QY 74 GGCCAGCTGCTCCCTGGTGAAGCGCAGGCTGGCTGGTGAATGAGAAAGTCTTCTCCGGA 133
 DB 912 GGACATGCTGGAGTGGAGGAGCGGAGGAGTGGTGGAGGAGCTGGCGGTGCCCAT 971

QY 134 GGCTGCTCCGCTGGTGCAGAGAGAGCTCAGGAGCTGGCAGAGTGGTGGCGGCTTGGAG 193
 DB 972 GGTGGTGGATACGCTGGTGGTGAAGAGTGGTGGTGGAGGAGGATATGGAGGTGCCGT 1031

QY 194 GCTGCTGGAAGAAAGTCTGCTGAGCCTCATCAGAACTGCGATCTGCAGAGGATGGAAGT 253
 DB 1032 GCAGGTGGATGAGGTGGTGGAGCGGTGGAAATGGAGCGGTGGAGGAGTGGAGGT 1091

QY 254 GGATTCGGGGAAGAAATGCT 274
 DB 1092 GCACAGGTGGTGGATACGCT 1112

RESULT 2
 US-08-781-891-208/c
 ; Sequence 208, Application US/08781891
 ; Patent No. 6090620
 ; GENERAL INFORMATION:
 ; APPLICANT: Fu, Ying-Hui
 ; APPLICANT: Yu, Chang-En
 ; APPLICANT: Oshima, Junko
 ; APPLICANT: Mulligan, John T.
 ; APPLICANT: Schellenberg, Gerald D.
 ; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 ; TITLE OF INVENTION: WERNER'S SYNDROME
 ; NUMBER OF SEQUENCES: 209
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; FILING DATE: 27-DEC-1996
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: No. 6090620tenburg Ph.D., Carol
 ; REGISTRATION NUMBER: 39,317
 ; REFERENCE/DOCKET NUMBER: 240052.419
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 208:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 16442 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-781-891-208

Query Match 10.1%; Score 49.8; DB 3; Length 16442;
 Best Local Similarity 49.4%; Pred. No. 0.0016;
 Matches 129; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 9 CCGGGGATCCGAGTCCCAAGAGCCGAGTTTGAGAGGCTGGTGGCAGAAATTCGCCGAG 68
 DB 16369 CAGGAGCAGGAGGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 16310

QY 69 AAGGAGGCCAGCTGTCTCCCTGGTGAAGCGCAGGCTGGCTGGTGAATGAGAAAGTCTTCT 128
 DB 16309 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 16250

QY 129 CCGGAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 188
 DB 16249 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 16190

QY 189 TTGAGCTGCTGGAGAAAGTCTGCTGAGCCTCATCAGAACTGCGCATCTGCAGAGGATG 248
 DB 16189 AAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 16130

QY 249 GAAGTGGATTCGGGGAGAA 269
 DB 16129 AAGAAGAAAGAAAGAAAGAA 16109

RESULT 3
 US-09-618-166-208/c
 ; Sequence 208, Application US/09618166
 ; Patent No. 6583112
 ; GENERAL INFORMATION:
 ; APPLICANT: Fu, Ying-Hui
 ; APPLICANT: Yu, Chang-En
 ; APPLICANT: Oshima, Junko
 ; APPLICANT: Mulligan, John T.
 ; APPLICANT: Schellenberg, Gerald D.
 ; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 ; TITLE OF INVENTION: WERNER'S SYNDROME
 ; NUMBER OF SEQUENCES: 209
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed Intellectual Property Law Group
 ; STREET: 701 Fifth Avenue, Suite 6300
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/618,166
 ; FILING DATE: 17-Jul-2000
 ; CLASSIFICATION: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mcmasters, David D.
 ; REGISTRATION NUMBER: 33,963
 ; REFERENCE/DOCKET NUMBER: 240052.419C1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 208:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 16442 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 208:
 ; US-09-618-166-208

Query Match 10.1%; Score 49.8; DB 4; Length 16442;
 Best Local Similarity 49.4%; Pred. No. 0.0016;
 Matches 129; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 9 CCGGGGATCCGAGTCCCAAGAGCCGAGTTTGAGAGGCTGGTGGCAGAAATTCGCCGAG 68
 DB 16369 CAGGAGCAGGAGGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 16310

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 13:24:31 ; Search time 436 Seconds
(without alignments)
6693.652 Million cell updates/sec

Title: US-10-071-510A-16

Perfect score: 493

Sequence: 1 cggcgccggcggggatgcc.....cctgtctcatttgagcctgc 493

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 16Dec04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2002bs:*

8: Geneseqn2003as:*

9: Geneseqn2003bs:*

10: Geneseqn2003cs:*

11: Geneseqn2003ds:*

12: Geneseqn2004as:*

13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	493	100.0	493	10	ADH69132 Human tum
2	347.6	70.5	4083	5	AAS83954 DNA encod
3	345.4	70.1	3003	12	ADQ67408 Novel hum
C 4	176.4	35.8	532	12	ACH77474 Human gen
C 5	174	35.3	174	12	ACH91174 Human gen
6	56.8	11.5	2282	8	AAL51693 Kukulcani
7	51.4	10.4	1505	2	AAQ55750 Genomic c
8	50.2	10.2	15231	6	ABK63598 Rat seque
9	50.2	10.2	15231	10	ADBS9205 Toxicity-
10	50.2	10.2	15231	10	ADBS3026 Primary r
11	50.2	10.2	15231	10	ABT42078 Toxicity
12	50.2	10.2	15231	12	ADP72968 Renal tox
13	49.8	10.1	768	6	ABZ13007 Arabidops
C 14	49.8	10.1	16442	2	AXX83006 Partial m
15	49.4	10.0	8298	4	AAK72613 Human imm
16	49	9.9	8973	12	ADP28653 Human sec
C 17	48.6	9.9	114793	4	AAO8215 Human gen
18	48.2	9.8	628	12	ACH9664 Human gen
19	47.8	9.7	799	2	AAV55831 Nucleotid
C 20	47.8	9.7	913	2	AAT07199 Cotton fi

C 21	47.8	9.7	913	2	AAT13034	Aat13034 Cotton fi
C 22	47.8	9.7	913	2	AAT30253	Aat30253 Cotton fi
C 23	47.8	9.7	913	2	AAT62610	Aat62610 Cotton fi
C 24	47.8	9.7	913	2	AAT70041	Aat70041 Cotton fi
C 25	47.8	9.7	913	3	AAZ35545	Aaz35545 cDNA sequ
26	47.8	9.7	1925	2	AAx90924	Aax90924 Epstein B
27	47.8	9.7	1926	3	AAAS0254	Aaas0254 Epstein B
28	47.8	9.7	1926	4	AAAF82902	Aaf82902 EBV tethe
C 29	47.8	9.7	1926	10	ADK65580	Adk65580 Human her
C 30	47.8	9.7	1984	2	AAT13030	Aat13030 Cotton fi
C 31	47.8	9.7	1985	2	AAT30250	Aat30250 Cotton fi
C 32	47.8	9.7	1985	2	AAT70036	Aat70036 Cotton H6
C 33	47.8	9.7	2580	3	AAAF75454	Aaaf75454 Nucleotid
C 34	47.8	9.7	2580	6	AAAF64275	Aaaf64275 Epstein-B
C 35	47.8	9.7	5452	2	AAAX90923	Aax90923 Anti-sens
C 36	47.8	9.7	8705	2	AAZ23778	Aaz23778 Vector ps
C 37	47.8	9.7	8705	12	ADM10659	Adm10659 Expressio
C 38	47.8	9.7	9482	12	ADP64415	Adp64415 Vector pc
C 39	47.8	9.7	9600	2	AAV21683	Aav21683 Vector pl
C 40	47.8	9.7	10285	6	ABS71037	Abs71037 PCBP-Xa-F
C 41	47.8	9.7	10285	6	ABS66453	Abs66453 Plasmid p
C 42	47.8	9.7	10330	12	ADL67154	Adl67154 Plasmid p
C 43	47.8	9.7	10380	2	AAZ22248	Aaz22248 Nucleotid
C 44	47.8	9.7	10477	12	ADL67152	Adl67152 Plasmid p
C 45	47.8	9.7	10516	12	ADL67150	Adl67150 Plasmid p

ALIGNMENTS

RESULT 1

ADH69132

ID ADH69132 standard; DNA; 493 BP.

AC ADH69132;

DT 25-MAR-2004 (first entry)

DE Human tumour resistance/sensitivity marker DNA #16.

XX human; ds; tumour; tumour growth; cancer; resistance; sensitivity.

XX Homo sapiens.

XX US2003143552-A1.

XX 31-JUL-2003.

XX 08-FEB-2002; 2002US-00071510.

XX 08-FEB-2001; 2001US-0267276P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Clark E, Grenfell-Lee T, Lu K, Hartmann L, Brown JL;

XX WPI; 2003-829783/77.

Determining whether agent can/cannot be used to reduce growth of tumor involves exposing obtained tumor cells to test agent, determining expression level of markers by tumor cells.

XX Disclosure; SEQ ID NO 16; 41pp; English.

XX The invention relates to a method of determining whether an agent can/cannot be used to reduce the growth of tumour, involves obtaining a sample of tumour cells, exposing them to one or more test agents, determining the level of expression of one or more markers by the cells exposed to the agent and by the cells not exposed to the agent and identifying the agent as appropriate/inappropriate to reduce the growth of tumour. The method is useful for determining whether an agent can/cannot be used to reduce the growth of tumour. The method is useful for determining whether an anti-cancer agent should in treatment of a

CC cancer patient should be continued/discontinued. The method is useful for
CC reducing growth rate of cancer in a patient. The present sequence
CC represents a human tumour resistance/sensitivity DNA marker.

XX
SQ Sequence 493 BP; 99 A; 124 C; 174 G; 96 T; 0 U; 0 Other;

Query Match 100.0%; Score 493; DB 10; Length 493;
Best Local Similarity 100.0%; Pred. No. 6.4e-114;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCGCGCCGCGGGATGCGAGTCCCAAGAGCCGAGTTTGAGAGCTGGTGGCAGAAAT 60

DB 1 CGCGCGCCGCGGGATGCGAGTCCCAAGAGCCGAGTTTGAGAGCTGGTGGCAGAAAT 60

QY 61 TCCCGAGAGAGGCGCCAGCTGCTCCCTGGTGGAGCGCAGGGCTGGCTGATGGAGA 120

DB 61 TCCCGAGAGAGGCGCCAGCTGCTCCCTGGTGGAGCGCAGGGCTGGCTGATGGAGA 120

QY 121 AGTCTTCTCCGAGGGTGTGCGGTGGTGCAGGAGGAGCTCAGGGAGCTGGCAGAGTCTGT 180

DB 121 AGTCTTCTCCGAGGGTGTGCGGTGGTGCAGGAGGAGCTCAGGGAGCTGGCAGAGTCTGT 180

QY 181 GCGCGCCCTTGAGGCTGTGCGAAGAAAGTCTGCTGAGCTCATCAGAAACTGGCATCTGC 240

DB 181 GCGCGCCCTTGAGGCTGTGCGAAGAAAGTCTGCTGAGCTCATCAGAAACTGGCATCTGC 240

QY 241 AGAGGATGAACTGGATTCCGGGAGAAATGGTTTCCACCAACATCCCAAGTCCAG 300

DB 241 AGAGGATGAACTGGATTCCGGGAGAAATGGTTTCCACCAACATCCCAAGTCCAG 300

QY 301 GATTTCTCATCATCCATCCATGGATCCTATTCCAGGAGCTGTCGACCGTGTGTCTTA 360

DB 301 GATTTCTCATCATCCATCCATGGATCCTATTCCAGGAGCTGTCGACCGTGTGTCTTA 360

QY 361 GCAGGCTGTGGAGAGGGGCGAGCCCGCAGCGTCAAGAGTGGGTAGGGGTCTCCAGCA 420

DB 361 GCAGGCTGTGGAGAGGGGCGAGCCCGCAGCGTCAAGAGTGGGTAGGGGTCTCCAGCA 420

QY 421 CAGGCGCCCTTCTGCTGGGCAACATCTGCTCTGAGGACTTGGCCAGCTCTGTCT 480

DB 421 CAGGCGCCCTTCTGCTGGGCAACATCTGCTCTGAGGACTTGGCCAGCTCTGTCT 480

QY 481 CATTTGAGCTGC 493

DB 481 CATTTGAGCTGC 493

RESULT 2

AAS83954

ID AAS83954 standard; cDNA; 4083 BP.

XX AAS83954;

AC AAS83954;

DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #19758.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

KW Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

DR P-PSDB; ABG19767.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

XX

PS Claim 1; SEQ ID NO 19758; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is

CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic

CC coding sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 4083 BP; 923 A; 1177 C; 1228 G; 755 T; 0 U; 0 Other;

Query Match 70.5%; Score 347.6; DB 5; Length 4083;

Best Local Similarity 98.9%; Pred. No. 3.7e-77;

Matches 350; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GCGCGCGCCGCGGGATGCGAGTCCCAAGAGCCGAGTTTGAGAGCTGGTGGCAGAAAT 61

DB 2796 GCGCGCGCCGCGGGATGCGAGTCCCAAGAGCCGAGTTTGAGAGCTGGTGGCAGAAAT 2855

QY 62 CCGGAGAGAGGAGGCGCCAGCTGCTCCCTGGTGAAGCGCAGGCTGGCTGATGAGAA 121

DB 2856 CCGGAGAGAGGAGGCGCCAGCTGCTCCCTGGTGAAGCGCAGGCTGGCTGATGAGAA 2915

QY 122 GTCTTCTCCGAGGGTGTGCGGTGGTGCAGGAGGAGCTCAGGGAGCTGGCAGAGTCGTG 181

DB 2916 GTCTTCTCCGAGGGTGTGCGGTGGTGCAGGAGGAGCTCAGGGAGCTGGCAGAGTCGTG 2975

QY 182 GCGGCGCTTGAGGCTGTGGAAAGAAAGTCTGTGAGCTCATCAGAAACTGGCATCTGCA 241

DB 2976 GCGGCGCTTGAGGCTGTGGAAAGAAAGTCTGTGAGCTCATCAGAAACTGGCATCTGCA 3035

QY 242 GAGGATGGAAGTGGATTCCGGGAGAAAGTGGTTTCCACCAACATCCCAAGTCAGS 301

DB 3036 GAGGATGGAAGTGGATTCCGGGAGAAAGTGGTTTCCACCAACATCCCAAGTCAGS 3095

QY 302 ATTTCTCATCAATCCCATGGATCCTATTCCAGGAGCTCGTCGACGCGTGTCT 355

DB 3096 ATTTCTCATCAATCCCATGGATCCTATTCCAGGAGCTCGTCGACGCGTGTCT 3149

RESULT 3

ADQ67408

ID ADQ67408 standard; cDNA; 3003 BP.

XX ADQ67408;

AC ADQ67408;

XX

DT 07-OCT-2004 (first entry)

XX

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 13:24:37 ; Search time 2364 Seconds
(without alignments)
10105.085 Million cell updates/sec

Title: US-10-071-510A-16
Perfect score: 493
Sequence: 1 cggcgccggccgggagatgcc.....cctgtctcatttgagcctgc 493

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	490.4	99.5	5502	9 HSM808450	BX648302 Homo sapi
2	347.6	70.5	1938	9 AK098471	AK098471 Homo sapi
3	347.6	70.5	3978	6 CQ735548	CQ735548 Sequence
4	345.4	70.1	3003	6 CQ845922	CQ845922 Sequence
5	345.4	70.1	3003	9 AK131436	AK131436 Homo sapi
6	283.4	57.5	150350	9 CNS01DWV	AL138539 Human chr
7	283.4	57.5	166007	9 CNS01DRC	AL117186 Human chr
8	121.4	24.6	176510	10 AC147375	AC147375 Mus muscu
9	57	11.6	142268	8 AP003018	AP003018 Oryza sat
10	57	11.6	154248	8 AP003631	AP003631 Oryza sat
11	57	11.6	159749	8 AP003020	AP003020 Oryza sat
12	56.8	11.5	2279	3 AY571308	AY571308 Kukulcani
13	56	11.4	154137	8 AP003215	AP003215 Oryza sat
14	55.4	11.2	135511	8 AP005097	AP005097 Oryza sat
15	55.4	11.2	154441	8 AP005383	AP005383 Oryza sat
16	55.2	11.2	194736	10 AC126271	AC126271 Mus muscu
17	54	11.0	109894	2 AP005975	AP005975 Oryza sat
18	54	11.0	132526	8 AP006451	AP006451 Oryza sat
19	54	11.0	147782	8 AP005604	AP005604 Oryza sat

20	53.8	10.9	55528	3 AC105764	AC105764 Caenorhab
21	53.6	10.9	78835	8 AC120528	AC120528 Oryza sat
22	53.4	10.8	169511	8 AP005309	AP005309 Oryza sat
23	53	10.8	541	8 CR354530	CR354530 Pinus pin
24	53	10.8	128223	8 AP005643	AP005643 Oryza sat
25	53	10.8	176530	8 AP003259	AP003259 Oryza sat
26	52.8	10.7	125020	9 AF429315	AF429315 Homo sapi
27	52.8	10.7	217522	10 AC117639	AC117639 Mus muscu
28	52.8	10.7	227897	2 AC111962	AC111962 Rattus no
29	52.8	10.7	237997	2 AC093989	AC093989 Rattus no
30	52.6	10.7	253504	10 AC102602	AC102602 Mus muscu
31	52.2	10.6	97839	2 AC141520	AC141520 Rattus no
32	52.2	10.6	151987	8 AP004683	AP004683 Oryza sat
33	52.2	10.6	199348	8 AC137751	AC137751 Oryza sat
34	52	10.5	146538	8 AC123594	AC123594 Oryza sat
35	52	10.5	300029	8 AE017081	AE017081 Oryza sat
36	51.8	10.5	215769	10 AC100209	AC100209 Mus muscu
37	51.8	10.5	244842	2 AC106417	AC106417 Rattus no
38	51.6	10.5	218236	2 AC108828	AC108828 Mus muscu
39	51.6	10.5	251872	2 AC125998	AC125998 Rattus no
40	51.4	10.4	1158	8 BNGRP22G	215045 B.napus GRP
41	51.4	10.4	137127	8 AC145323	AC145323 Oryza sat
42	51.4	10.4	162311	8 AC145321	AC145321 Oryza sat
43	51.4	10.4	191753	10 AC125235	AC125235 Mus muscu
44	51.4	10.4	248750	10 AC112685	AC112685 Mus muscu
45	51.2	10.4	280	8 CR377334	CR377334 Pinus pin

ALIGNMENTS

RESULT 1
HSM808450
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
CONSRM
TITLE
JOURNAL
COMMENT

HSM808450
Homo sapiens mRNA; cDNA DKFp686M07125 (from clone DKFp686M07125).
BX648302.1
GI:34367461

Homo sapiens (human)

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 5502)

Bloeker, H., Boecker, M., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,

Fobo, G., Han, M. and Wiemann, S.

The German Human cDNA Consortium

Direct Submission

Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764

Neuherberg, GERMANY

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by GFP (National Research Centre for Biotechnology Ltd.,

Braunschweig/Germany) within the cDNA sequencing consortium of the

German Genome Project.

This clone (DKFp686M07125) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further

information about the clone and the sequencing project is available

at http://mips.gsf.de/proj/cDNA/.

FEATURES

Location/Qualifiers

1..5502

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFp686M07125"

/tissue_type="human uterus"

/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host

DH10B; sites SfiIA + SfiIB"

/dev_stage="adult"

5474

polyA_site

ORIGIN

Query Match

99.5%; Score 490.4; DB 9; Length 5502;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 9, 2005, 20:31:26 ; Search time 86 Seconds
(without alignments)
5871.047 Million cell updates/sec

Title: US-10-071-510A-16

Perfect score: 923

Sequence: 1 cggccggccgggggagtcg.....cctgtctcattgagcctgc 493

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=xlip
-Q=/cgn2_1/USPTO.spool_p/US10071510/runat_09022005_132509_18476/app_query.fasta_1.647
-DB=UniProt 03 -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pcpt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10071510.cgn_1_152@runat_09022005_132509_18476 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 03.*

1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	592	64.1	489	2 Q8N7G8	Q8N7G8 homo sapien
2	586	63.5	904	2 Q6ZM23	Q6ZM23 homo sapien
3	404	43.8	895	2 Q8C117	Q8C117 mus musculus
4	305	33.0	262	2 Q8BMM1	Q8BMM1 mus musculus
5	149	16.5	555	1 G91 CHLRE	Q9FPQ6 chlamydomon
6	141	15.6	1143	2 Q948Y6	Q948Y6 volvox cart
7	140.5	15.6	3889	2 Q6SSE8	Q6SSE8 chlamydomon
8	139.5	15.5	3409	2 Q6SSE6	Q6SSE6 chlamydomon
9	138	15.3	636	2 Q7SF15	Q7SF15 neurospora
10	138	15.3	676	2 Q95JC9	Q95JC9 sus scrofa
11	136.5	15.1	745	2 Q89X06	Q89X06 bradyrhizob
12	136	15.1	236	2 Q6QJ26	Q6QJ26 arabidopsis
13	136	15.1	342	2 Q6ZD62	Q6ZD62 oryza sativ
14	136	15.1	731	2 Q65530	Q65530 arabidopsis
15	135	15.0	659	2 Q6C7Q8	Q6C7Q8 yarrowia li
16	134.5	14.9	464	2 Q41645	Q41645 volvox cart

c 17	134.5	14.9	566	2	Q95JD1	Q95jd1 sus scrofa
c 18	134	14.9	511	2	Q95UD0	Q95ud0 sus scrofa
c 19	133.5	14.8	309	2	Q18751	Q18751 caenorhabdi
c 20	133.5	14.8	687	2	Q948Y7	Q948y7 volvox cart
c 21	133.5	14.8	1997	2	Q8LRM7	Q8lrm7 chlamydomon
c 22	132.5	14.7	698	2	Q9ASK4	Q9ask4 oryza sativ
c 23	131.5	14.6	326	2	Q22514	Q22514 santalum al
c 24	131.5	14.6	763	2	Q9XDH2	Q9xdh2 mycobacteri
c 25	131.5	14.6	1033	1	IF2 STRCO	Q6cjg8 streptomyce
c 26	131.5	14.6	3204	2	Q6X248	Q6x248 bovine herp
c 27	130.5	14.5	708	2	Q9SX31	Q9sx31 arabidopsis
c 28	130.5	14.5	708	2	Q82327	Q82327 arabidopsis
c 29	130	14.4	191	2	Q9C580	Q9c580 arabidopsis
c 30	130	14.4	525	2	Q95L53	Q95l53 mustela vis
c 31	129.5	14.4	386	2	Q9FPQ5	Q9fpq5 chlamydomon
c 32	129	14.3	839	2	Q9SN46	Q9sn46 arabidopsis
c 33	129	14.3	1146	2	Q6PLP6	Q6plp6 chlamydomon
c 34	129	14.3	1334	2	Q9RKR9	Q9rkr9 streptomyce
c 35	128.5	14.3	598	2	Q8VKR7	Q8vkn7 mycobacteri
c 36	128	14.2	409	2	Q9SBM1	Q9sbm1 volvox cart
c 37	128	14.2	430	2	Q40385	Q40385 nicotiana a
c 38	127.5	14.2	1213	2	Q92923	Q92923 homo sapien
c 39	127.5	14.2	1533	2	Q7SAT8	Q7sat8 neurospora
c 40	127	14.1	448	2	Q9WPM2	Q9wpm2 macaca mula
c 41	127	14.1	471	2	Q6AXT8	Q6axt8 rattus norv
c 42	127	14.1	475	1	S3A2 MOUSE	Q72n23 mus musculu
c 43	127	14.1	485	2	Q7TN25	Q7tn25 mus musculu
c 44	127	14.1	551	2	Q16630	Q16630 homo sapien
c 45	126.5	14.0	464	1	S3A2_HUMAN	Q15428 homo sapien

ALIGNMENTS

RESULT 1

Q8N7G8 PRELIMINARY; PRT; 489 AA.
ID Q8N7G8
AC Q8N7G8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein FLJ25605.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Suzuki Y., Hata H.,
RA Nakagawa K., Mizuno S., Morinaga M., Kawamura M., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Nishikawa T., Sugano S.;
RA Nagai K., Isegai T., Sugano S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK098471; BAC05312.1; -
DR Genew; HGNC:19861; C14orf49.
DR InterPro; IPR02017; Spectrin.
DR SMART; SM00150; SPEG; 1.
SQ SEQUENCE 489 AA; 56176 MW; 19ABD8DE00ACE6EF CRC64;

Alignment Scores:
Pred No.: 2.44e-38 Length: 489
Score: 592.00 Matches: 117
Percent Similarity: 95.16% Conservative: 1
Best Local Similarity: 94.35% Mismatches: 6
Query Match: 64.14% Indels: 0
DB: 2 Gaps: 0

US-10-071-510A-16 (1-493) x Q8N7G8 (1-489)

Qy 3 GCCGCCCGGGGATCGGAGTCCAGAGCCCGAGCTTGAGAGCTGTCGACAAATTC 62

```

Db 192 AlaGlyProGlyAspAlaGluSerGlnGluAlaGluPheGluArgLeuValAlaGluPhe 211
QY 63 CCGGAGAAGGAGCCAGCTGCTCCCTGGTGGAGCCAGCGCTGGCTGGTGGAGAG 122
Db 212 ProGluGlyGluAlaGlnLeuSerLeuValGluAlaGlnGlyTrpLeuValMetGluLys 231
QY 123 TCTTCTCCGAGGAGCTGCTCCGCTGGTGGAGGAGCTCAGGAGCTGGCAGAGTCTGTG 182
Db 232 SerSerProGluGlyAlaValAlaValGlnGluLeuArgGluLeuAlaGluSerTrp 251
QY 183 CCGGCTTGGAGTCTGGAAGAAAGTCTGAGCTCATCAGAACTGGCATCTGCAG 242
Db 252 ArgAlaLeuArgLeuGluSerLeuLeuSerLeuLeuSerLeuLeuSerLeuGln 271
QY 243 AGGATGGAAGTGGATTCCGGGAGAGAAATGTTTCCACCAACACATCCCAAGTCAGGA 302
Db 272 ArgMetGluValAspSerGlyLysLysMetValPheThrAsnAsnLeuProLysSerGly 291
QY 303 TTTCTCATCAATCCCATGATCTATTCCAGGAGCTGTCGAGCGTGGTCTGTAGC 362
Db 292 PheLeuLeuAsnProMetAspProLeuProArgHisArgArgAlaAsnLeuGln 311
QY 363 AGGCTGTGGGA 374
Db 312 GluGluGluGly 315

RESULT 2
Q6ZM23
ID Q6ZM23 PRELIMINARY; PRT; 904 AA.
AC Q6ZM23;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Hypothetical protein FLJ16564.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Synovial membrane tissue;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Iehii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK131436; BAD18582.1; -.
DR InterPro: IPR002017; Spectrin.
DR Pfam: PF00435; Spectrin; 1.
DR SMART: SM00150; SPEC; 2.
SQ SEQUENCE 904 AA; 104056 MW; CA4C7405B17CACC9 CRC64;

Alignment Scores:
Pred. No.: 7,91e-38 Length: 904
Score: 586.00 Matches: 115
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 63.49% Indels: 0
DB: 2 Gaps: 0

```

US-10-071-510A-16 (1-493) x Q6ZM23 (1-904)

```

QY 3 GCGGCGCCGCGGAGTCCGAGTCCCAAGAGCCGAGTTTGAGAGGCTGGTGGCAGAAATTC 62
Db 612 AlaGlyProGlyAspAlaGluSerGlnGluAlaGluPheGluArgLeuValAlaGluPhe 631
QY 63 CCGGAGAAGGAGCCAGCTGCTCCCTGGTGGAGCCAGCGCTGGCTGGTGGAGAG 122
Db 632 ProGluLysGluAlaGlnLeuSerLeuValGluAlaGlnGlyTrpLeuValMetGluLys 651

```

```

QY 123 TCTTCTCCGAGGAGTCTCCCTGGTGGAGGAGCTCAGGAGCTGCAGAGTCTGTGG 182
Db 652 SerSerProGluGlyAlaValAlaValGlnGluLeuArgGluLeuAlaGluSerTrp 671
QY 183 CCGGCTTGGAGTCTGGAAGAAAGTCTGAGCTCATCAGAACTGGCATCTGCAG 242
Db 672 ArgAlaLeuArgLeuGluSerLeuLeuSerLeuLeuSerLeuLeuSerLeuGln 691
QY 243 AGGATGGAAGTGGATTCCGGGAGAGAAATGTTTCCACCAACACATCCCAAGTCAGGA 302
Db 692 ArgMetGluValAspSerGlyLysLysMetValPheThrAsnAsnLeuProLysSerGly 711
QY 303 TTTCTCATCAATCCCATGATCTATTCCAGGAGCTGTCGAGCGC 347
Db 712 PheLeuLeuAsnProMetAspProLeuProArgHisArgArg 726

RESULT 3
Q8C117
ID Q8C117 PRELIMINARY; PRT; 895 AA.
AC Q8C117;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched
DE library, clone:4831426I19 product:hypothetical Spectrin repeat
DE containing protein, full insert sequence.
GN Name=4831426I19Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RC The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Suganara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Washiwagi K.,
RA Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanishi M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 9, 2005, 21:36:36 ; Search time 26.5 Seconds
(without alignments)

3579.990 Million cell updates/sec

Title: US-10-071-510A-16

Perfect score: 923

Sequence: 1 cggcggccgggggagtcg.....cctgtctcatttgagcctgc 493

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2 1/USPTO.spool_p/US10071510/runat 09022005_132509_18488/app_query.fasta_1.647
-DB=PIR_79 -OPMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10071510.cgn 1 1 38 @runat 09022005_132509_18488 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_79:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	136	15.1	731	T04455	hypothetical prote
C 2	134.5	14.9	464	S22697	extensin - Volvox
C 3	133.5	14.8	309	T32933	hypothetical prote
C 4	130.5	14.5	708	D96711	hypothetical prote
C 5	130	14.4	191	F84522	probable proline-r
C 6	129	14.3	839	T04859	extensin homolog F
C 7	129	14.3	1334	T50568	probable multi-dom
C 8	127	14.1	551	S57447	HPBRII-7 protein -
C 9	126.5	14.0	464	A47655	spliceosome-associ
C 10	122.5	13.6	283	S13383	hydroxyproline-ric
C 11	121.5	13.5	1006	T42731	atrophin-1 related
C 12	121	13.4	222	H96711	hypothetical prote
C 13	121	13.4	929	C96623	hypothetical prote
C 14	120.5	13.4	1392	T51947	probable transcrip

C 15 120 13.3 760 2 F86387
C 16 119.5 13.3 3149 1 QQBEB8
C 17 119 13.2 339 2 T2607
C 18 119 13.2 461 2 T10741
C 19 118.5 13.2 388 2 JC5437
C 20 118.5 13.2 1736 2 T00391
C 21 118 13.1 224 2 D72861
C 22 118 13.1 439 2 S51939
C 23 118 13.1 534 2 S21961
C 24 118 13.1 1015 2 JC6552
C 25 117.5 12.7 313 2 A28444
C 26 117.5 13.0 515 2 F70904
C 27 117.5 13.0 576 2 T36729
C 28 117 13.0 227 2 C29149
C 29 117 13.0 907 2 E96636
C 30 117 13.0 1188 2 S49915
C 31 117 13.0 2715 2 T13049
C 32 116.5 12.9 240 2 A24264
C 33 116.5 12.9 291 2 T20942
C 34 116.5 12.9 329 2 T32783
C 35 116 12.9 240 2 B24264
C 36 116 12.9 317 2 A28996
C 37 116 12.9 377 2 A48018
C 38 116 12.9 533 2 S37781
C 39 116 12.9 599 2 T10798
C 40 116 12.9 1137 2 A86335
C 41 115.5 12.8 214 2 T10737
C 42 115.5 12.8 214 2 T09854
C 43 115.5 12.8 265 2 T46089
C 44 115.5 12.8 306 2 T52340
C 45 115.5 12.8 620 2 S06733

ALIGNMENTS

RESULT 1

T04455

hypothetical protein F4D11.90 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 16-Aug-2004

C:Accession: T04455

R:Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Hoheisel, J.; Mewes, H. W.

submitted to the Protein Sequence Database, April 1998

A:Reference number: Z15360

A:Accession: T04455

A:Molecule type: DNA

A:Residues: 1-731 <BEV>

A:Cross-references: UNIPROT:O65530; EMBL:AL022537

A:Experimental source: cultivar Columbia; BAC clone F4D11

C:Genetics:

A:Map position: 4

A:Intron: 326/1; 476/1; 505/1; 528/3; 557/2; 606/3; 660/3

A:Note: F4D11.90

C:Superfamily: protein kinase homology

Alignment Scores:
Pred. No.: 0.00256 Length: 731
Score: 136.00 Matches: 52
Percent Similarity: 33.52% Conservative: 8
Best Local Similarity: 29.05% Mismatches: 54
Query Match: 15.09% Indels: 65
DB: 2 Gaps: 6

US-10-071-510A-16 (1-493) x T04455 (1-731)

QY 442 CCCCCACAGAGGGGCGCTGTCTGAGACCCCTACCC-----AC 401

10 AlaProLaThrSerProAlaMetSerLeuPro-ProAlaAspSerValProAspTh 29

QY 400 CTCTTGACCTGGGCGCTGGCCCTTCTCCACAGCCCTGCTAGACAGCTCAGCGTCGA 341

29 rSerSerProAlaProLeuSerProLeuPro----- 41

```
QY 340 CGATGCTGGGAATAGATCCATGGATTGATGAGAAATCTGACTTTGGGATTTGTTG 281
Db 41 ----- 41
QY 280 GTGAAACCATTTCTTCCCGAATCCATCTTCAGATGCCAGTTTCTGATG 221
Db 42 ----ProProLeuSerSerProProProLeuProSerProProProLeuSerAlaProTh 60
QY 220 AGGCTCAGCAGACTTCTTCCAGCAGCCCTCAAGCCCGCCAGCAGCTCTGCCAGCTCC-- 163
Db 60 rAlaSerProProProLeuProValGluSerProProSerProProProLeuGluSerProPr 80
QY 162 -----TGAGCTCTCTCTGCACCCAGCGCAGCACCT-- 133
Db 80 oProProLeuLeuGluSerProProProProProProProLeuGluSerProProSerPr 100
QY 132 -----CCGAGAAGACTTCTCC 116
Db 100 oHisValSerAlaProSerGlySerProProLeuProPheLeuProAlaLysProSerPr 120
QY 115 ATCAGCAGCCAGCCCT-----GGCTTCCACAGGAGCAGCTGGCGCTCTCTCTCC 65
Db 120 oProProSerSerProProSerGluThrValProProGlyAsnThrIleSerProProPr 140
QY 64 GGGAATTTGCCACCCAGCCCTCAAACTCGGCTCTTGGGACTCGGCATCCCCCG 10
Db 140 oArgSerLeuProSerGluSer---ThrProProValAsnThrAlaSerProPro 157
RESULT 2
S22697
extension - Volvox carteri (fragment)
C;Species: Volvox carteri
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Jul-2004
C;Accession: S22697; S21006
R;Ertl, H.; Hallmann, A.; Wenzl, S.; Sumper, M.
EMBO J. 11, 2055-2062, 1992
A;Title: A novel extension that may organize extracellular matrix biogenesis in Volvox ca
A;Reference number: S22697; MUID:92289669; PMID:1600938
A;Accession: S22697
A;Molecule type: mRNA
A;Residues: 1-464 <HAL>
A;Cross-references: UNIPROT:Q41645; EMBL:X65165; NID:g21991; PIDN:CAA46283.1; PID:g21992
C;Keywords: glycoprotein
Alignment Scores:
Pred. No.: 0.0033 Length: 464
Score: 134.50 Matches: 40
Percent Similarity: 40.68% Conservative: 8
Best Local Similarity: 33.90% Mismatches: 33
Query Match: 14.93% Indels: 37
DB: 2 Gaps: 5
US-10-071-510A-16 (1-493) x S22697 (1-464)
QY 273 CCATTTTCTTCCCGAATCCATCTCCATCTCTGCAGATGCCAGTTTCTGATGAGGTCA 214
Db 325 ProArgSerSerProSerProProProProProProProProProProProProProProPro 339
QY 213 GCAGACTTTCTTCCAGCAGCTCAAGGCCCGCCAGCAGCTCTGCCAGCTCCCTGAGCT--- 157
Db 340 -----SerProProProProProProSerProProProProProProProProProSer 356
QY 156 -----CCTCTGCACACCGGAGCACCTCCG 130
Db 357 ProSerProProProProValValSerProProProProProProProProProProProPro 376
QY 129 GAGAAGACTTCTCCATCAGCAGCCAGCTGGCTTCCACAGGAGCAGCTGGCGCTCTCT 70
Db 377 ProProProAlaSerSerProProProProProProProProProProProProProProPro 396
QY 69 TCTCCGGGAATTTGCCACCGCC-- 281
Db 397 SerProPro-----ProProAlaThrAlaAlaAlaAsnProProSerProProAlaProSer 414
```

```
QY 42 CAAACTCGG-----CCTCTGGGACTCGCATCCCCCGGCGCGCG 1
Db 415 ArgSerArgAlaGlyGlyProProLeuGlyThrArgProProProProProPro 432
RESULT 3
T29293
hypothetical protein C50F7.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T29293
R;Johnson, D.; Stellyes, L.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid C50F7.
A;Reference number: Z20601
A;Accession: T29293
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-309 <JOH>
A;Cross-references: UNIPROT:Q18751; EMBL:U41557; PIDN:AAA83301.1; CESP:C50F7.5
C;Genetics:
A;Gene: CESP:C50F7.5
Alignment Scores:
Pred. No.: 0.00387 Length: 309
Score: 133.50 Matches: 52
Percent Similarity: 36.81% Conservative: 8
Best Local Similarity: 31.90% Mismatches: 62
Query Match: 14.82% Indels: 42
DB: 2 Gaps: 7
US-10-071-510A-16 (1-493) x T29293 (1-309)
QY 467 CCAGTCTCTCAGACGACAGATGTTGCCCCAGACAGGAGGGGCGCTGTCTGGAGACCCC 408
Db 151 ProGlyProProValAspProSerGluAspProGlnProSerValGluProSerGluAspPro 170
QY 407 TACCCACCTCTTGACCTGGCGCTGCGCCCTCTTCCACAGCCCTCTAGACAGACTCAC 348
Db 171 GlnProSerGlyProProSerProGlyProValAspProSerGluAspProGlnProSer 190
QY 347 GCGTCAGCAGCTCGCTGGGAATAG-GATCCATGGGATGTAGAGAAATCCTGACTTTGGGA 289
Db 191 GlySerSerSerProGlyProValAspProSerAsp----- 202
QY 288 TGTGTGTGTGAAACCATTTCTTCCCGAATCTTCCATCTCTGCAGATGCCAGT 229
Db 203 -----GluProSerProSerProSerProSerProGlyProValAspProSer 218
QY 228 TTCTGATGAGGCTCAGCAGACTTTCTTCCAGCAGCCTCAAGGCCCGCCAGCAGCTCTGCCA 169
Db 219 -----GluAspPro 221
QY 168 GTCTCCCTGAGCTCTCTCCACACCGCAGCACCTCCGAGAGAGACTTCTCCATCACCACCA 109
Db 222 LysProSerGluProProSerProGlyProValAspPro---SerAspGluProSerPro 240
QY 108 GCAGCCCTCGGCTTCCACAGGAGCAGCTGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 58
Db 241 SerAspProProGlyProProGlyProProGlyProProGlyProProGlyProProGlyPro 259
QY 57 CTGCCACAGCCCTCTCAAACTCGG-----CCTCTTGGG-----ACTCGGCAT 16
Db 260 ---ProProGlyProProThrArgArgProProGlyProProGlyProProThrArgArg 278
QY 15 CCCCCGGGC 7
Db 279 ProProGly 281
```

```
RESULT 4
D96711
hypothetical protein F24J5.8 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 10, 2005, 03:01:00 ; Search time 505.5 Seconds
(without alignments)
637.339 Million cell updates/sec

Title: US-10-071-510A-16

Perfect score: 923

Sequence: 1 cggcgccggggggatgcc.....cctgtctcattgagcctgc 493

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 2753750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10071510/runat_09022005_132511_18595/app_query.fasta_1.647
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bite -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DLOCAL=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10071510@cgn_1_130@runat_09022005_132511_18595
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pcp.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pcp.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
C 1	149.5	16.6	220	16	US-10-437-963-139610	Sequence 139610,
C 2	148	16.4	263	16	US-10-437-963-116465	Sequence 116465,
C 3	144.5	16.0	437	16	US-10-437-963-168762	Sequence 168762,
C 4	143	15.9	148	16	US-10-437-963-189874	Sequence 189874,
C 5	142	15.8	501	16	US-10-437-963-161137	Sequence 161137,
C 6	138	15.3	155	16	US-10-767-701-40674	Sequence 40674, A
C 7	138	15.3	189	16	US-10-767-701-36014	Sequence 36014, A
C 8	138	15.3	639	15	US-10-369-493-3962	Sequence 3962, Ap
C 9	137.5	15.1	313	16	US-10-437-963-135163	Sequence 135163,
C 10	136.5	15.1	360	16	US-10-437-963-179473	Sequence 179473,
C 11	136	15.1	378	16	US-10-437-963-119242	Sequence 119242,
C 12	136	15.1	547	16	US-10-437-963-102579	Sequence 102579,
C 13	136	15.1	731	13	US-10-086-464-17	Sequence 17, Appl
C 14	135	15.0	447	16	US-10-437-963-152072	Sequence 152072,
C 15	134.5	14.9	270	16	US-10-437-963-130068	Sequence 130068,
C 16	134.5	14.9	380	16	US-10-437-963-200034	Sequence 200034,
C 17	134	14.9	187	16	US-10-767-701-35648	Sequence 35648, A
C 18	134	14.9	285	16	US-10-437-963-139612	Sequence 139612,
C 19	134	14.9	533	16	US-10-437-963-176617	Sequence 176617,
C 20	133.5	14.8	299	16	US-10-437-963-134054	Sequence 134054,
C 21	133.5	14.8	442	16	US-10-437-963-197867	Sequence 197867,
C 22	132.5	14.7	273	16	US-10-437-963-184760	Sequence 184760,
C 23	132.5	14.7	756	15	US-10-424-599-200688	Sequence 200688,
C 24	132.5	14.7	760	15	US-10-425-114-39701	Sequence 39701, A
C 25	132	14.7	130	15	US-10-424-599-250148	Sequence 250148,
C 26	132	14.7	373	16	US-10-437-963-149274	Sequence 149274,
C 27	131.5	14.6	317	16	US-10-437-963-143427	Sequence 143427,
C 28	130.5	14.5	191	16	US-10-767-701-31817	Sequence 31817, A
C 29	130.5	14.5	200	16	US-10-437-963-134619	Sequence 134619,
C 30	130.5	14.5	240	16	US-10-437-963-160874	Sequence 160874,
C 31	130	14.4	201	16	US-10-437-963-174274	Sequence 174274,
C 32	130	14.4	270	16	US-10-437-963-197222	Sequence 197222,
C 33	130	14.4	272	16	US-10-437-963-197111	Sequence 197111,
C 34	130	14.4	418	16	US-10-437-963-142252	Sequence 142252,
C 35	130	14.4	563	16	US-10-437-963-198755	Sequence 198755,
C 36	129.5	14.4	141	16	US-10-437-963-152071	Sequence 152071,
C 37	129.5	14.4	313	16	US-10-437-963-114711	Sequence 114711,
C 38	129	14.3	431	16	US-10-437-963-204963	Sequence 204963,
C 39	129	14.3	554	16	US-10-437-963-119177	Sequence 119177,
C 40	129	14.3	555	16	US-10-437-963-203630	Sequence 203630,
C 41	128.5	14.3	134	16	US-10-437-963-137023	Sequence 137023,
C 42	128.5	14.3	193	16	US-10-767-701-57745	Sequence 57745, A
C 43	128.5	14.3	306	16	US-10-437-963-146544	Sequence 146544,
C 44	128.5	14.3	344	16	US-10-437-963-173746	Sequence 173746,
C 45	128.5	14.3	358	16	US-10-437-963-105278	Sequence 105278,

ALIGNMENTS

RESULT 1
US-10-437-963-139610
; Sequence 139610, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 139610
; LENGTH: 220
; TYPE: PRT

; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(220)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_40885C.1.pap
US-10-437-963-139610

Alignment Scores:
Pred. No.: 0.00094 Length: 220
Score: 149.50 Matches: 47
Percent Similarity: 33.54% Conservative: 8
Best Local Similarity: 28.66% Mismatches: 58
Query Match: 16.59% Indels: 51
DB: 16 Gaps: 6

US-10-071-510A-16 (1-493) x US-10-437-963-139610 (1-220)

QY 417 TGGAGACCCCTACCCACCTTTGACCTGGGGCTGGCCCTTCTC----- 373
DB 76 TrpCysProProleuserProProCysProTrpProTrpTrp***CysProPro 95
QY 372 -----CCACAGCCCTGTAGACAGACTCACGGCTCGACGATGCCTGGGAATAGGAT 322
DB 96 ProProProProGlnProProProProGlnProProProProTrpSerProGlu 115
QY 321 CCATGGGATTGATGAGAAATCCTGACTTTGGGATGTTGT-----TGGTGAAAA 274
DB 116 LysTrp-----CysProProProTrpSerPro 125
QY 273 CCATT-----TCCTCC 262
DB 126 ProTrpTrpCysGlnProProPro***SerProProCysSerTrpCysProProSer 145
QY 261 CCGAATTCACCTTCCTCTGCAGATGCCAGTTTCTGATGAGGCTCAGCAGACTTTCTT 202
DB 146 ProProProSerProProSerGlyProSerSer----- 156
QY 201 CCAGCAGCCTCAAGGCCCGCCACGACTCTGCCAGCTCCTCTGACCTCTCTGCACACGG 142
DB 157 ProProProSerProPro-----ProProProSerAlaProProProSer 172
QY 141 CAGCACCTCCGGAGAAGACTTCTCATCACGCCAGCCCTCGCGTTTCCACGAGGACA 82
DB 173 AlaProProProGlnProSerAlaProProProSerProProSerProProPro 192
QY 81 GCTGGCCCTCTTCTCCGGGAATTCTGCCACAGCCTCTCAAACTCGGCTCTTGGGACT 22
DB 193 ProProProProSerProProProSerProProAlaProGlnProSerLeuAenLeu 212
QY 21 CGGCATCCCCCG 10
DB 213 AlaValAlaPro 216

RESULT 2

US-10-437-963-116465
; Sequence 116465, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 116465
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(263)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_19964C.1.pap
US-10-437-963-116465

Alignment Scores:
Pred. No.: 0.00128 Length: 263
Score: 148.00 Matches: 43
Percent Similarity: 36.17% Conservative: 8
Best Local Similarity: 30.50% Mismatches: 42
Query Match: 16.43% Indels: 48
DB: 16 Gaps: 5

US-10-071-510A-16 (1-493) x US-10-437-963-116465 (1-263)

QY 411 CCCCTACCCACCTTTGACCTGGGGCTGGCCCTTCTCCACAGCCCTGTAGACAGAC 352
DB 62 ProAlaProThrProProProAla---ProThrThrProProProAlaProThr--- 79
QY 351 TCACGGCTCGACGATGCTGGGAATAGGATCCATGGGATTGATGAGAAATCCTGACTTTG 292
DB 79 ----- 79
QY 291 GGATGTTGTTGGTGAACCAATTTCTCCCGAATCCACTTCCATCTTCGAGATGCC 232
DB 80 -----ThrProProProAlaProThrThr--- 87
QY 231 AGTTTCTGATGAGGCTCAGCAGACTTTCTTCCACGAGCCCTCAAGGCCCGCCACGACTCTG 172
DB 88 -----ProProProSerProProAlaThrProPro 97
QY 171 CCAGCTCCCTGAGCTCTCTCT-----GCACACGGCAGCACCCTCCGGAGAGACTTCT 118
DB 98 ProAlaProThrThrProProProSerProProSerGlnProProProAlaProAlaThr 117
QY 117 CCATCACCAGCAGCCCTCGGCTTCCACGAGGACAGCTGGG-----CCTCCTTCTCCG 64
DB 118 ProProProSerProProAlaThrProProProAlaProAlaThrProProProSerPro 137
QY 63 GGAATTCGCCACCCAGCCTCTCAAACTCGGCTCTTGGGACTCGGCATCCCCGGGCCGG 4
DB 138 ProLeuAlaProProProAlaThrProProProProAlaThrProProAla***Arg 157
QY 3 CCG 1
DB 158 Pro 158

RESULT 3

US-10-437-963-168762
; Sequence 168762, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 10, 2005, 02:47:34 ; Search time 30 Seconds
(without alignments)

2453.466 Million cell updates/sec

Title: US-10-071-510A-16

Perfect score: 923

Sequence: 1 cggccggccgggggatgc.....cctgtctcatttgagcctgc 493

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10071510/runat_09022005_132510_18506/app_query.fasta_1.647
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10071510@cgn 1_1_33 @runat_09022005_132510_18506 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	130	14.4	420	US-09-902-540-13993	Sequence 13993, A
C 2	127	14.1	553	US-09-949-016-7961	Sequence 7961, Ap
C 3	126.5	14.0	324	US-09-949-016-7664	Sequence 7664, Ap
C 4	126	13.7	339	US-09-252-991A-21715	Sequence 21715, A
5	125.5	13.6	561	US-09-252-991A-22317	Sequence 22317, A
C 6	123	13.7	334	5202236-3	Patent No. 5202236
C 7	123	13.7	334	5202236-3	Patent No. 5202236
C 8	122.5	13.6	802	US-09-823-240A-2	Sequence 2, Appli
C 9	122	13.5	331	5202236-37	Patent No. 5202236
C 10	122	13.5	331	5202236-37	Patent No. 5202236
C 11	120.5	13.4	581	US-09-949-016-9978	Sequence 9978, Ap
C 12	120	13.3	174	US-08-818-112-143	Sequence 143, App

C 13	120	13.3	174	3	US-08-818-111-138	Sequence 138, App
C 14	120	13.3	174	3	US-09-056-556-143	Sequence 143, App
C 15	120	13.3	174	3	US-09-072-596-138	Sequence 138, App
C 16	120	13.3	174	4	US-09-072-967-143	Sequence 143, App
C 17	120	13.0	701	4	US-09-252-991A-24048	Sequence 24048, A
C 18	119	13.2	525	3	US-08-764-870-7	Sequence 7, Appli
C 19	119	13.2	525	3	US-08-980-115-7	Sequence 7, Appli
C 20	119	13.2	533	1	US-07-952-800-2	Sequence 2, Appli
C 21	119	13.2	533	4	US-08-216-592A-4	Sequence 4, Appli
C 22	119	13.2	577	4	US-09-949-016-11572	Sequence 11572, A
C 23	118	13.1	447	4	US-09-252-991A-22113	Sequence 22113, A
C 24	118	13.1	511	4	US-09-107-433-2904	Sequence 2304, Ap
C 25	117	13.0	104	4	US-09-547-693-235	Sequence 235, App
C 26	116	12.9	180	4	US-09-504-615-78	Sequence 78, Appli
C 27	115.5	12.8	214	1	US-08-217-327-4	Sequence 4, Appli
C 28	115.5	12.5	356	4	US-09-252-991A-18206	Sequence 18206, A
C 29	115.5	12.5	445	4	US-09-252-991A-23505	Sequence 23505, A
C 30	115	12.8	267	3	US-08-301-162-16	Sequence 16, Appl
C 31	115	12.8	267	3	US-09-461-240-16	Sequence 16, Appl
C 32	115	12.8	267	4	US-09-968-927-16	Sequence 16, Appl
C 33	115	12.8	398	3	US-09-303-064-54	Sequence 54, Appl
C 34	115	12.8	398	3	US-09-086-503-54	Sequence 54, Appl
C 35	114.5	12.7	105	4	US-09-547-693-230	Sequence 230, App
C 36	114	12.7	72	4	US-09-547-693-231	Sequence 231, App
C 37	113.5	12.6	503	4	US-09-599-287A-2	Sequence 2, Appli
C 38	112.5	12.2	226	4	US-09-252-991A-29689	Sequence 29689, A
C 39	112.5	12.2	1958	1	US-07-945-283-2	Sequence 2, Appli
C 40	112.5	12.5	8991	4	US-08-714-741-32	Sequence 32, Appl
C 41	111.5	12.1	606	3	US-08-556-978B-21	Sequence 21, Appl
C 42	111.5	12.1	606	3	US-09-247-806-4	Sequence 4, Appli
C 43	111	12.3	174	4	US-09-248-796A-25211	Sequence 25211, A
C 44	111	12.3	316	4	US-09-252-991A-25345	Sequence 25345, A
C 45	111	12.0	577	4	US-09-252-991A-31927	Sequence 31927, A

ALIGNMENTS

RESULT 1

US-09-902-540-13993
; Sequence 13993, Application US/0902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 13993
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-13993

Alignment Scores:			
Pred. No.:	0.000404	Length:	420
Score:	130.00	Matches:	42
Percent Similarity:	36.57%	Conservative:	7
Best Local Similarity:	31.34%	Mismatches:	35
Query Match:	14.43%	Indels:	50
DB:	4	Gaps:	7

US-10-071-510A-16 (1-493) x US-09-902-540-13993 (1-420)

QY	276	AAACCAATTTCTTC	-----CGG 259
Db	105	LysprophGiuSerGlnValLeuLeuAspLysValLysAlaLeuValGlyClnLysSer	124

QY 258 AATCCACTTCCATCTCTGCAGATGCCAGTTTCTGATGAGGCTCAGCAGACTTCTTCTCCA 199
||| :|||:|||||
Db 125 AsnThrMetProAlaSerAla-----AlaThrGlnValArgHis 137
||| :|||:|||||
QY 198 GCAGCTCAAGGCCGCCAGCACTTGCAGCTTCCC-----TGAGCTCTCTCTGCA 148
||| :|||:|||||
Db 138 AlaAlaProGlnProAlaAlaAla---ProAlaProValAlaAlaAlaProProGly 156
||| :|||:|||||
QY 147 CCAGCGCAGCACCCTCGGAGAGACTTCTCCATCA-----112
||| :|||:|||||
Db 157 AlaArgProAlaProProGlyAlaArgProGlyValProProGlyProGlyValPro 176
||| :|||:|||||
QY 111 -----CCAGCCAGCCCTGGCTTCCA 91
||| :|||:|||||
Db 177 ArgProProGlyAlaGlyValProProProGlyAlaArgProProGlyProGlyMetPro 196
||| :|||:|||||
QY 90 CCAGGACAGCTGGGCTCTCTTCCGGGAATTCCTCCACCACTCT-----CAAACT 37
||| :|||:|||||
Db 197 ProGlyMetAlaArgProProGlyProGlyMetProProGlyAlaProGlyAlaPro 216
||| :|||:|||||
QY 36 CGGCTCTTGGGACTCGGCATCCCGGGC-----CGGCG 1
||| :|||:|||||
Db 217 ArgProProGlyProGlyMetProProGlyMetAlaArgPro 230
||| :|||:|||||
RESULT 2
US-09-949-016-7961
; Sequence 7961, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7961
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7961
Alignment Scores:
Pred. No.: 0.000855 Length: 553
Score: 127.00 Matches: 67
Percent Similarity: 35.92% Conservative: 7
Best Local Similarity: 32.52% Mismatches: 77
Query Match: 14.10% Indels: 56
Gaps: 4
US-10-071-510A-16 (1-493) x US-09-949-016-7961 (1-553)
QY 467 CCAAGTCTCAGAGCAGCATGTGCCCCAGACAGGGA-----429
||| :|||:|||||
Db 191 ProGlyGlySerSerArgAlaAlaPheProGlnGlyGlyArgGlyArgPhePro 210
||| :|||:|||||
QY 428 GGGGCTGTGCTGAGACCCCTACCCACTCTTTGACCTGGGGCTGGC-----381
||| :|||:|||||
Db 211 GlyAlaValProGlyGlyArgPhePheProGlyProAlaGlyProGlyGlyProPro 230
||| :|||:|||||
QY 380 CCTTCTCCACAGCCCTGTAGACAGACTCAGCGTCGACGATGCTGGGA---ATAGG 324
||| :|||:|||||
Db 231 ProPhe-----ProAlaGlyGlnThr-ProProArgProProLeuGlyProProGly 247
||| :|||:|||||
QY 323 ATCCATGGGATTGATGAGAAATCTGACTTTGGGATGTTCTGTGTAACCACTTTCTT 264
||| :|||:|||||

Db 247 yProProGlyProProGlyProProProGlyGlnValLeuPro-ProProLeuAlaG 267
||| :|||:|||||
QY 263 CCCCC-----AATCCACTTCCATCTCTGCAGATGCCAGTTTCTGATGAG 219
||| :|||:|||||
Db 267 lyProProAsnArgGlyAspArgProProPro-ProValLeuPheProGlyGlnProPhe 286
||| :|||:|||||
QY 218 GCTCAGCAGACTTCTTCCAGCAGCTCAAGGCCGCCACG-ACTCTGCCAGCT-----166
||| :|||:|||||
Db 287 GlyGlnProProLeuProProGlyProProProProProValProGlyTyrgly 306
||| :|||:|||||
QY 165 ---CCTAGCTCTCTCTGCACACGGCAGCACCTCCGAGAGAACTCTTCATCACC 109
||| :|||:|||||
Db 307 ProProProGlyProProProGlnGlnGlyProProProProGlyProPhePro 326
||| :|||:|||||
QY 108 GCCAGCCCT-----GGCTTCCACCGGACAGCT 79
||| :|||:|||||
Db 327 ProArgProProGlyProLeuGlyProProLeuThrLeuAlaProProHisLeuPro 346
||| :|||:|||||
QY 78 GGGCTCTCTTCTCCGGGAATTCGCCACCACTCTCAAACTCGGCTCTT-----28
||| :|||:|||||
Db 347 GlyProPro---ProGlyAlaProProProAlaProHisValAsnProAlaPhePro 365
||| :|||:|||||
QY 27 -----GGACTCGGCATCCCCCG-----10
||| :|||:|||||
Db 366 ProProThrAsnSerGlyMetProThrSerAspSerArgGlyProProProThrAspPro 385
||| :|||:|||||
QY 9 ---GGCGGCGC 1
||| :|||:|||||
Db 386 TyrglyArgPro 389
||| :|||:|||||
RESULT 3
US-09-949-016-7664
; Sequence 7664, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7664
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7664
Alignment Scores:
Pred. No.: 0.000807 Length: 324
Score: 126.50 Matches: 52
Percent Similarity: 38.79% Conservative: 12
Best Local Similarity: 31.52% Mismatches: 66
Query Match: 14.04% Indels: 35
Gaps: 4
US-10-071-510A-16 (1-493) x US-09-949-016-7664 (1-324)
QY 440 CCAGACAGAGGGGGCTGTGCTGGAGACCCCTACCCACTCTTGACCTGGGGCTGGC 381
||| :|||:|||||
Db 125 ProProThrGlyProAlaProSerGlyProProGlyProGlnLeuProProAla 144
||| :|||:|||||
QY 380 CCTTCTCCACAGCCCTGTAGACAGACTCAGCGTCGACGATGCTGGGAATGATC 321
||| :|||:|||||
Db 145 ProGlyValHisProProAlaProValHisPro-----ProAlaSerGlyVal 161
||| :|||:|||||

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 9, 2005, 19:39:31; Search time 88 seconds
(without alignments)

4333.474 Million cell updates/sec

Title: US-10-071-510A-16

Perfect score: 923

Sequence: 1 cggccggccggggatgcc.....cctgtctcatttgagcctgc 493

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=xlp
-Q=/cgn2 1/USPTO.spool p/US10071510/runat 09022005.132508.18468/app_query.fasta_1.647
-DB=A Geneseq_16Dec04 -OPMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pcio -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10071510@cgn 1_154@runat_09022005.132508.18468 -NCPUS=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_16Dec04:
1: Geneseqp1980s:
2: Geneseqp1990s:
3: Geneseqp2000s:
4: Geneseqp2001s:
5: Geneseqp2002s:
6: Geneseqp2003as:
7: Geneseqp2003bs:
8: Geneseqp2004s:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	586	63.5	904	ADQ67715	ADQ67715 Novel hum
2	156.5	17.4	325	ABG21919	ABG21919 Novel hum
3	138	15.3	639	ADN21309	ADN21309 Bacterial
4	136	15.1	731	AAW74209	AAW74209 Protein e
5	136	15.1	731	ABR93202	ABR93202 Herbicida
6	136	15.1	731	ABR82942	ABR82942 Arabidops
7	133	14.8	406	ABG27250	ABG27250 Novel hum
8	133	14.8	598	ABG14000	ABG14000 Novel hum
9	133	14.8	598	ADP60423	ADP60423 Human con
10	131.5	14.6	763	AAW31852	AAW31852 Mycobacte

C 11	130.5	14.5	708	5	ABB91504	Abb91504 Herbicida
C 12	130	14.4	191	3	AAG29836	Aag29836 Arabidops
C 13	130	14.4	199	3	AAG29835	Aag29835 Arabidops
C 14	128.5	14.3	598	6	ABU37030	Abu37030 Protein e
C 15	127	14.1	399	3	AAB43375	Aab43375 Human ORF
C 16	127	14.1	448	3	AAB53201	Aab53201 Macaca mu
C 17	126.5	14.0	259	8	ABO58334	AbO58334 Human gen
C 18	126.5	14.0	464	6	ABO52933	AbO52933 Human spl
C 19	126.5	14.0	464	8	ABM81998	AbM81998 Tumour-as
C 20	126	13.7	339	7	ABO72969	AbO72969 Pseudomon
C 21	126	14.0	800	8	ABO58564	AbO58564 Human gen
C 22	125.5	13.6	561	7	ABO73571	AbO73571 Pseudomon
C 23	125	13.9	205	6	ABU20105	AbU20105 Protein e
C 24	124.5	13.8	647	4	AAW74205	Aaw74205 Brassica
C 25	124.5	13.8	647	7	ABR82937	AbR82937 B. napus
C 26	124.5	13.8	728	7	ABR82938	AbR82938 B. napus
C 27	123	13.7	334	1	AAW82971	Aaw82971 Biocadhesi
C 28	123	13.7	510	2	AAW39469	Aaw39469 hXR-beta
C 29	123	13.7	572	2	AAW31855	Aaw31855 Mycobacte
C 30	123	13.7	1023	5	AAU82954	Aau82954 Human hom
C 31	122.5	13.6	783	2	AAW37151	Aaw37151 Mouse neu
C 32	122.5	13.6	787	2	AAW37152	Aaw37152 Mouse neu
C 33	122.5	13.6	802	2	AAW37153	Aaw37153 Mouse neu
C 34	122.5	13.6	802	4	AAU09139	Aau09139 Mammalian
C 35	122	13.5	334	1	AAW83194	Aaw83194 Sequence
C 36	122	13.5	415	4	ABG30150	AbG30150 Novel hum
C 37	122	13.5	434	3	AAW58798	Aaw58798 Breast an
C 38	121.5	13.5	116	4	AAW20321	Aaw20321 Peptide #
C 39	121.5	13.5	116	4	ABB40803	Abb40803 Peptide #
C 40	121.5	13.5	116	4	AAW34569	Aaw34569 Peptide #
C 41	121.5	13.5	116	4	ABB24992	Abb24992 Protein #
C 42	121.5	13.5	116	4	AAW74455	Aaw74455 Human bon
C 43	121.5	13.5	116	4	AAW61662	Aaw61662 Human liv
C 44	121.5	13.5	116	4	ABG56248	AbG56248 Human liv
C 45	121.5	13.5	116	5	ABG44340	AbG44340 Human pep

ALIGNMENTS

RESULT 1

ADQ67715

ID ADQ67715 standard; protein; 904 AA.

XX ADQ67715;

XX AC ADQ67715;

XX DT 07-OCT-2004 (first entry)

XX DE Novel human protein sequence #2381.

XX DE Novel human protein sequence #2381.

XX DE osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;

XX DE gene therapy; diagnostic marker; morbid state; osteoporosis;

XX DE neurological disease; Alzheimer's disease; Parkinson's disease; dementia;

XX DE cancer.

XX OS Homo sapiens.

XX OS Homo sapiens.

XX PN EPI440981-A2.

XX XX 28-JUL-2004.

XX XX 21-JAN-2004; 2004EP-00001196.

XX XX 21-JAN-2003; 2003JP-00102206.

XX XX 09-MAY-2003; 2003JP-00131392.

XX XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

XX XX Yamamoto J, Isono Y, Nagai K, Irie R;

XX XX WPI; 2004-535376/52.

XX XX N-PSDB; ADQ67408.

PT Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
 PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
 XX
 PS Claim 1; SEQ ID NO 4876; 2449pp; English.
 XX
 CC The invention relates to 2495 novel polynucleotides (I) and their encoded
 CC polypeptides, sequences hybridizing to these nucleotides, sequences
 CC encoding partial polypeptides and sequences having 70% or 90% identity to
 CC the nucleotide and protein sequences. The nucleotides and polypeptides
 CC are useful as diagnostic markers or therapeutic target for the diseases
 CC or morbid states. They are also useful for treating osteoporosis,
 CC neurological diseases, Alzheimer's diseases, Parkinson's diseases,
 CC dementia and various cancers. This sequence corresponds to a protein
 CC sequence of the invention.
 XX
 SQ Sequence 904 AA;

Alignment Scores:
 Pred. No.: 1,21e-48 Length: 904
 Score: 586.00 Matches: 115
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.49% Indels: 0
 DB: 8 Gaps: 0

US-10-071-510A-16 (1-493) x ADQ67715 (1-904)

QY 3 GCCGCGCCGGGATCCGAGTCCCAAGAGCCGAGTTTGAGAGGTGGGCGAGATTC 62
 DB 612 AlaGlyProGlyAspAlaGluSerGlnGluAlaGluPheGluValAlaGluPhe 631
 QY 63 CCGAGAGAGGAGCCAGCTGCTCCGTGGAGCCGAGGCTGGGTGGTGGAGAG 122
 DB 632 ProGluGysGluAlaGlnLeuSerLeuValGluAlaGlnGlyTrpLeuValMetGluLys 651
 QY 123 TCTTCTCCGAGGGTGTGCGGTGGTGGAGAGGCTCAGGAGGCTGGCAGAGTCTGTGG 182
 DB 652 SerSerProGluGlyAlaAlaValGlnGluLeuArgGluLeuAlaGluSerTrp 671
 QY 183 CGGGCTTGAGGTGCTGGAGAAAGTCTGCTGAGGCTCATCAGAACTGGCATCTGCGAG 242
 DB 672 ArgAlaLeuArgLeuLeuGluGluSerLeuLeuSerLeuIleArgAsnTrpHisLeuGln 691
 QY 243 AGCATGGAGTGGATTCGGGGAGAGAAATGGTTTTCACCAACACATCCCAAGTCAAGGA 302
 DB 692 ArgMetGluValAspSerGlyLysLysMetValPheThrAsnAsnIleProLysSerGly 711
 QY 303 TTTCTCATCAATCCATGGATCCTATTCCCGAGGATCGTCCGAGCC 347
 DB 712 PheLeuIleAsnProMetAspProIleProArgHisArgArgArg 726

RESULT 2
 ID ABG21919
 ID ABG21919 standard; protein; 325 AA.

AC ABG21919;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #21910.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.

XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.
 DR N-PSDB; AAS86106.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 52278; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 325 AA;

Alignment Scores:
 Pred. No.: 2,39e-06 Length: 325
 Score: 156.50 Matches: 47
 Percent Similarity: 38.41% Conservative: 6
 Best Local Similarity: 34.06% Mismatches: 34
 Query Match: 17.37% Indels: 51
 DB: 4 Gaps: 6

US-10-071-510A-16 (1-493) x ABG21919 (1-325)

QY 408 CTACCCACCTCTTGACCTGGGCGCTGCTCCAGAGCCCTGCTAGACAGACTCA 349
 DB 101 LeuProProProProProAlaSerProProLeuLeuPro----- 113
 QY 348 CGCGTCGACGATGCTGGGAATAGGATCCATGGATTGATGAGAAATCCTGACTTTGGGA 289
 DB 113 ----- 113
 QY 288 TGTTGTTGGTGAACACCATTTTCTCCCGAATCCACTTCCATCTCTCGAGATGCCAGT 229
 DB 114 -----ProAlaLeuSerPro---ProLeuProAlaProPro----- 124
 QY 228 TTCTGATGAGCTCAGCAGACTTTCTCCAGCAGCCTCAGGCGCCCGCCAGCTCTGCCA 169
 DB 125 -----ProProSerAlaProProAlaSerProProPro 135
 QY 168 GCTCCCTGAGCTCTCTGCACCGCAGCCCTCCGAGAGAACTTCTCCATCA--- 112
 DB 136 AlaProProProProProAlaProProProProProProProProProProAla 155
 QY 111 CCAGCCAGCCCTGCG---CTTCCACGAGGAGACAGTGGGCTCTCTCTCGGGAAATCTG 55
 DB 156 ProSerSerProAlaProLeuProProAlaProAlaSerProProSerProAlaProPro 175